

TESIS

**ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19
PADA FACEBOOK DAN INSTAGRAM MENGGUNAKAN
SUPPORT VECTOR MACHINE**



Disusun oleh:

Nama : Iln Kurniasari
NIM : 19.52.1213
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

TESIS

**ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19
PADA FACEBOOK DAN INSTAGRAM MENGGUNAKAN
SUPPORT VECTOR MACHINE**

**SENTIMENT ANALYSIS OF PUBLIC OPINION COVID-19
ON FACEBOOK AND INSTAGRAM
USING SUPPORT VECTOR MACHINE**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Iln Kurniasari
NIM : 19.52.1213
Konsentrasi : Informaties Technopneurshlp

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

HALAMAN PENGESAHAN

**ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19
PADA FACEBOOK DAN INSTAGRAM MENGGUNAKAN
SUPPORT VECTOR MACHINE**

**SENTIMENT ANALYSIS OF PUBLIC OPINION COVID-19
ON FACEBOOK AND INSTAGRAM
USING SUPPORT VECTOR MACHINE**

Dipersiapkan dan Disusun oleh

Iin Kurniasari

19.52.1213

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 6 Januari 2021

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 6 Januari 2021

Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN
ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19
PADA FACEBOOK DAN INSTAGRAM MENGGUNAKAN
SUPPORT VECTOR MACHINE

SENTIMENT ANALYSIS OF PUBLIC OPINION COVID-19
ON FACEBOOK AND INSTAGRAM
USING SUPPORT VECTOR MACHINE

Dipersiapkan dan Disusun oleh

Iin Kurniasari

19.52.1213

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 6 Januari 2021

Pembimbing Utama

Dr. Kusriani, M.Kom
NIK. 190302106

Pembimbing Pendamping

Hanif Al Fatta, M.Kom
NIK. 190302096

Anggota Tim Penguji

Prof. Ema Utami, S.Si, M.Kom
NIK. 190302037

Dr. Arief Sevanto, S.Si, MT
NIK. 190302036

Dr. Kusriani, M.Kom
NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 6 Januari 2021
Direktur Program Pascasarjana

Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Iin Kurniasari
NIM : 19.52.1213
Konsentrasi : Informatics Technopreneurship

Menyatakan bahwa Tesis dengan judul berikut:
ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA
FACEBOOK DAN INSTAGRAM MENGGUNAKAN SUPPORT VECTOR
MACHINE

Dosen Pembimbing Utama : Dr. Kusriani, M.Kom
Dosen Pembimbing Pendamping : Hanif Al Fatta, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

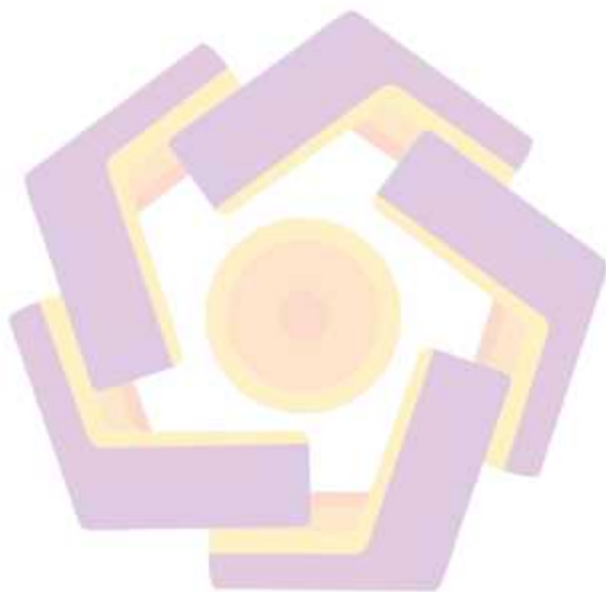
Yogyakarta, 6 Januari 2021
Yang Menyatakan,



Iin Kurniasari

HALAMAN PERSEMBAHAN

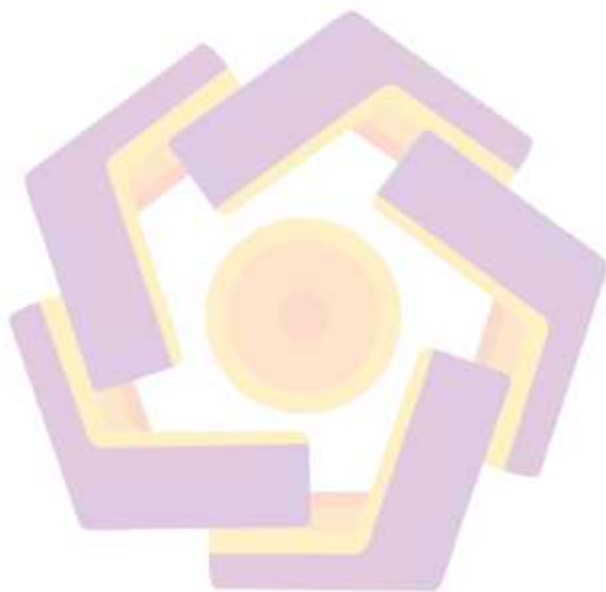
Tesis ini saya persembahkan kepada Allah SWT atas limpahan Rahmat-Nya, Guru, Orang Tua, Keluarga, sahabat dan semuanya yang telah memberikan dukungan moril dan spiritual. Semoga karya yang sedikit ini dapat memberikan manfaat dan kontribusi kedepannya. Amiin Ya Robbal 'Alamin.



HALAMAN MOTTO

Belajar **Berpikir Positif** dalam setiap keadaan karena:

“Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu; Allah mengetahui, sedang kamu tidak mengetahui.” (QS. Al-Baqarah: 216).



KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat, taufik serta hidayahNya sehingga penulis berhasil menyelesaikan tesis yang berjudul “ Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Facebook dan Instagram Menggunakan Support Vector Machine ”. Dalam hal ini penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Kusriani, M.Kom dan Bapak Hanif Al Fatta. M.Kom yang dengan sabarnya membimbing penulis dalam penelitian ini
2. Prof. Ema Utami, S.Si, M.Kom dan Bapak Dr. Arief Setyanto, S.Si, MT selaku dekan Peguji
3. Keluarga Besar Admisi MTI Amikom
4. Bapak Yudho Bismo Utomo, M.Kom, Ibu Aprilia Dian Evasari, MM serta keluarga besar Teknik Komputer Universitas Islam Kadiri
5. Kedua Orang Tua di Kediri dan di Malang beserta keluarga besar yang selalu mendoakan dalam setiap langkah penulis
6. Suami dan putra/ putri penulis, yang selalu memberikan dukungan dengan sepenuh hati.
7. Keluarga Besar di Bantul yang berkenan menyediakan tempat tinggal penulis selama belajar di Amikom Yogyakarta.
8. Serta seluruh pihak terkait yang tidak bisa saya sebutkan satu persatu.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, ibarat pepatah tak ada gading yang tak retak. Penulis memohon maaf jika ada kesalahan atau ketidaksesuaian, semoga tesis ini dapat memberikan kontribusi dan kebermanfaatannya untuk kedepannya.

Yogyakarta, Januari 2021

Penulis

DAFTAR ISI

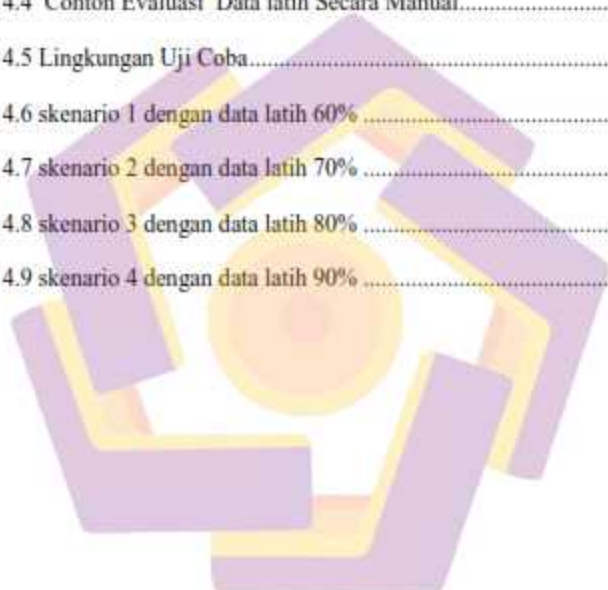
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR DIAGRAM.....	xiii
INTISARI.....	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1. Tinjauan Pustaka.....	8
2.2. Keaslian Penelitian.....	10
2.3. Landasan Teori.....	13

BAB III METODE PENELITIAN.....	25
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	25
3.2 Metode Pengumpulan Data.....	26
3.3 Metode Analisis Data.....	27
3.4 Alur Penelitian.....	27
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	29
4.1 Tahap Scraping Data Komentar Facebook dan Instagram.....	29
4.2 Pre Processing Facebook dan Instagram.....	33
4.3 Klasifikasi Support Vector Machine.....	38
BAB V PENUTUP.....	51
5.1 Kesimpulan.....	51
5.2. Saran.....	52
DAFTAR PUSTAKA.....	53



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian.....	9
Tabel 4.1 Contoh komentar pada Facebook dan Instagram AG243	31
Tabel 4.2 Proses Case Folding.....	32
Tabel 4.3 Pembagian jumlah Data Latih dan Data Uji	36
Tabel 4.4 Contoh Evaluasi Data latih Secara Manual.....	37
Tabel 4.5 Lingkungan Uji Coba.....	37
Tabel 4.6 skenario 1 dengan data latih 60%	39
Tabel 4.7 skenario 2 dengan data latih 70%	40
Tabel 4.8 skenario 3 dengan data latih 80%	41
Tabel 4.9 skenario 4 dengan data latih 90%	41

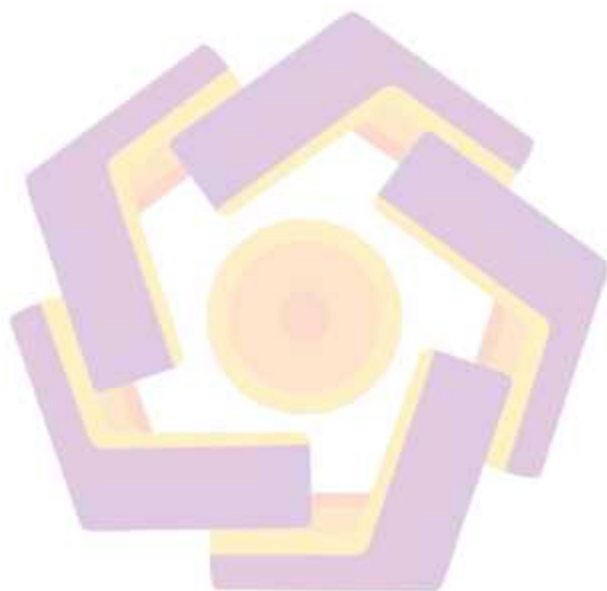


DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses KDD (Han dan Kamber 2006)	13
Gambar 2.2 Support Vector Machine	21
Gambar 2.3 Formula untuk Support Vector Machine.....	22
Gambar 2.4 Tampilan WebHarvy	23
Gambar 3.1 Alur Pengumpulan Data.....	25
Gambar 3.2 Alur Penelitian.....	27
Gambar 4.1 Tampilan Facebook pada webharvy.....	28
Gambar 4.2 Tampilan Instagram pada Webharvy	29
Gambar 4.3 Proses Scraping data FB dan IG pada webharvy	30
Gambar 4.4 Hasil Scraping Data.....	30
Gambar 4.5 Script Python untuk case folding	32
Gambar 4.6 Script tokenizing	33
Gambar 4.7 Contoh hasil Tokenizing	33
Gambar 4.8 Script filtering atau stopword removal pada Python.....	34
Gambar 4.9 Proses Filtering atau Stopword Removal.....	34
Gambar 4.10 Proses Stemming	35
Gambar 4.11 Script stemming pada Python.....	35

DAFTAR DIAGRAM

Diagram 4.1 Evaluasi Facebook	42
Diagram 4.2 Evaluasi Instagram	43



INTISARI

Perkembangan teknologi dewasa ini mendorong masyarakat untuk selalu tanggap teknologi, terlebih di era pandemi covid-19 yang selalu mengedepankan *social distancing*. Media sosial digunakan sebagai suatu alat untuk menyampaikan opini masyarakat kepada khalayak. Dalam penelitian ini, penulis melakukan penelitian tentang opini masyarakat pada media sosial facebook dan instagram dengan menggunakan *Support Vector Machine*.

Setelah dilakukan uji akurasi dan presisi ternyata SVM belum sesuai digunakan sebagai algoritma yang dapat menangkap urutan karena susunan kata yang dibolak balik meskipun maknanya berbeda tetap bermakna sama oleh mesin SVM, hal ini dibuktikan juga dengan jumlah akurasi yang kecil, yaitu 57% untuk facebook dan 59% untuk instagram. Sehingga diperlukan langkah untuk bisa diteliti dengan algoritma lain misalnya algoritma HRRN (Highest Response Ratio Next) atau LSTM (Long Short-Term Memory) yang memperhatikan urutan dan proses dengan rasio respon paling tinggi. Jika berdasarkan pendekatan ekstraksi fitur SVM dengan pendekatan count vector, tf-idf word level, tf-idf ngram level dan tf-idf char level. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur count vector dan tf-idf ngram level.

Kata kunci: analisis sentimen, instagram, covid-19, SVM, media sosial



ABSTRACT

Today's technological developments encourage people to always be responsive to technology, especially in the era of the Covid-19 pandemic which always prioritizes social distancing. Social media is used as a tool to convey public opinion to audiences. In this study, the authors conducted research on public opinion on the social media Facebook and Instagram using the Support Vector Machine.

After testing the accuracy and precision, it turns out that SVM is not suitable for use as an algorithm that can capture sequences because the word order is reversed even though the meaning is different, it still means the same by the SVM machine, this is also proven by the small amount of accuracy, namely 57% for Facebook and 59 % for instagram. So that steps are needed to be investigated with other algorithms such as the HRRN (Highest Response Ratio Next) algorithm or LSTM (Long Short-Term Memory) which takes into account the sequence and process with the highest response ratio. If based on the SVM feature extraction approach with the count vector approach, tf-idf word level, tf-idf ngram level and tf-idf char level. In this scenario the highest accuracy value is found in calculations using the count vector feature extraction and the tf-idf ngram level.

Keyword: sentiment analysis, instagram, covid-19, SVM, social media



BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi digital yang sangat pesat saat ini mendorong masyarakat harus tanggap teknologi digital, tidak hanya masyarakat elit perkotaan atau pelaku usaha. Masyarakat pinggiranpun juga harus tanggap sehingga tidak tertindas oleh era digital itu sendiri. Di era digital komunikasi tidak harus dilakukan dengan tatap muka, terlebih pada masa Pandemi Covid-19 yang mengedepankan *social distancing*. Hal ini dijadikan salah satu langkah pemerintah dalam menanggulangi penyebaran Covid-19 dengan memanfaatkan teknologi digital berupa media sosial dalam upaya *social distancing*.

Covid-19 merupakan sebuah ujian, ujian penguasa semesta terhadap penduduk dunia. Covid-19 bermula di Kota Wuhan China akhir 2019 dan kini telah menjangkiti lebih dari 100 negara. Lebih dari 100.000 orang di dunia dinyatakan positif Covid-19. Beberapa kasus Covid-19 di China dilaporkan menurun, tetapi lonjakan terjadi di Korea Selatan, Italia, Iran, Indonesia dan beberapa Negara lain. (Ikhsan: 2020)

Berdasarkan survei dari peta persebaran covid-19, di Indonesia ada lebih dari 399 kasus. 4241 terkonfirmasi, 3509 dalam perawatan, sembuh 359 dan meninggal 373. Sedangkan di Kediri (Kota dan Kabupaten) per 12 April 2020 di laman <http://infocovid19.jatimprov.go.id/> dilaporkan 9599 ODR (Orang Sehat dalam Resiko), 304 OTG (Orang Tanpa Gejala) teridentifikasi, 424 ODP (Orang

Dalam Pemantauan), 30 PDP (Pasien Dalam Pantauan) dan 17 Orang Positif Covid-19.

Media sosial digunakan sebagai media dalam penyampaian informasi khususnya dalam masa pandemi Covid-19 yang mengedepankan *social distancing*. Di Kediri media yang aktif dalam menginformasikan Covid-19 adalah AG243 karena beberapa media informasi formal yang dimiliki Satgas Covid tidak berfungsi secara maksimal dalam menampilkan informasi. Masyarakat banyak mengakses AG243 yang merupakan akun resmi yang dimiliki oleh Radio Andika FM yang terletak di Jl. Semeru 243 Kediri. Hal ini diakui oleh Waliota Kediri dalam Harian Radar Kediri 22 Mei 2020 tentang ucapan terima kasih bahwa situs AG243 selalu lebih dominan dalam membantu pemerintah menyampaikan informasi tentang covid-19.

AG adalah symbol Plat Nomor Kendaraan Kediri sedangkan 243 adalah nomor alamat Radio Andika. Dalam kesehariannya Radio Andika FM yang terhubung dengan Radio Suara Surabaya (SS) selalu menginformasikan kondisi keadaan jalan raya di area Karisedanan Kediri (Kediri, Madiun, Tulungagung, Trenggalek, Blitar dan sekitarnya).

AG 243 memiliki pengikut facebook 497.554, dan 78.300 pengikut IG dengan 12.000 postingan. Hampir setiap jam update info selalu dilakukan. Sehingga tidak heran dengan masa sekian banyak, setiap kali update informasi selalu banyak yang berkomentar. Khususnya dalam pandemi Covid-19 saat ini. Begitu banyak komentar baik positif, negatif ataupun netral yang diberikan khalayak terhadap update informasi yang disampaikan.

Dalam penelitian ini, penulis sangat tertarik menggunakan algoritma *Support Vector Machine*. Hal ini mengacu pada penelitian Wanda (2018) tentang analisis sentiment *cyberbullying* pada komentar instagram dengan menggunakan *Support Vector Machine (SVM)* dimana mendapatkan tingkat akurasi 90%. Tingkat akurasi yang cukup tinggi.

Akbar (2018) dalam penelitiannya tentang penerapan metode *Support Vector Machine (SVM)* dalam media sosial twitter dan facebook AMIKOM Yogyakarta menunjukkan facebook memiliki presentase lebih tinggi daripada twitter. Sedangkan Abu (2018) dalam penelitiannya tentang analisis sentiment komentar facebook dengan *K-Nearest Neighbor (KNN)* pada jasa ekspedisi barang J&T Express diperoleh hasil akurasi tertinggi 79,21 dan akurasi terendah 70,3 dengan jumlah dokumen training 101.

Eko (2019) dalam penelitiannya tentang analisis sentiment calon presiden Indonesia 2019 berdasarkan komentar public di facebook dengan menggunakan *Naïve Bayes* dengan hasil Joko Widodo 4242 komentar positif dan 758 komentar negatif. Sedangkan Prabowo Subianto memperoleh 3824 komentar positif dan 1176 komentar negatif.

Moehammad (2020) dalam penelitiannya tentang analisis sentiment tweet di Arab Saudi tentang pencegahan Pemerintah terhadap Covid-19. Moehammad menggunakan *Naïve Bayes* dalam penelitiannya dan dihasilkan bahwa kicauan di tweeter yang diteliti belum mewakili populasi masyarakat Arab Saudi karena banyak kicauan yang mengesampingkan tanda pagar sehingga tidak masuk dalam mesin analisis yang digunakan dan diperlukan survey di media sosial lain serta

dengan metode lain. James (2020) dalam penelitiannya yang berjudul analisis sentiment di media sosial dengan metode maximum entropy dihasilkan bahwa diperlukan kontribusi manusia dalam maximum entropy sehingga nilai akurasi tinggi.

AG243 memiliki 3 media sosial, 2 streaming yang aktif, dan dalam setiap status yang diunggah AG243 selalu mendapatkan *feedback* atau komentar yang sangat banyak. Dalam penelitian ini, penulis ingin mengetahui bagaimana opini masyarakat Kediri di media sosial AG243 radio Andika dalam menyebarkan informasi khususnya tentang Covid-19 serta penulis tertarik dengan algoritma Support Vector Machine untuk melakukan proses klasifikasi dengan nilai akurasi yang sangat signifikan yaitu 90% berdasarkan penelitian Wanda (2018). Tetapi berdasarkan penelitian Krusnoslav, dkk (2013) dalam *Comparison of Naïve Bayes and SVM Classifiers in Categorization of Concept Maps*. SVM hanya memiliki tingkat akurasi 63,74%. Sehingga berdasarkan fenomena di atas, penulis tertarik meneliti dan berinisiatif mengangkat judul “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Facebook dan Instagram Menggunakan Support Vector Machine” dalam penelitiannya.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka permasalahan yang akan dibahas dalam tesis ini dirumuskan sebagai berikut:

1. Pada media sosial manakah algoritma *Support Vector Machine* menghasilkan nilai akurasi tertinggi?

2. Fitur apa saja yang mempengaruhi nilai akurasi tertinggi pada metode *Support Vector Machine* pada *facebook* dan *instagram* ?

1.3. Batasan Masalah

Batasan masalah yang ditentukan untuk menghindari perluasan pembahasan dalam penelitian ini adalah sebagai berikut:

1. Data yang diklasifikasi adalah data komentar yang berisi informasi Covid-19 dengan kata kunci "covid19, covid-19, corona, covid dan pandemi" pada *facebook* dan *instagram* AG243 (radio andika).
2. Data komentar diambil dengan metode *scraping* data pada tanggal 18 Juli 2020 s/d 18 Agustus 2020 menggunakan *webharvy*
3. Data dianalisis dengan 4 (empat) tahap *preprocessing* yaitu: *casefolding*, *tokenizing*, *filtering*, *stemming* serta analisis sentiment menggunakan metode *Support Vector Machine (SVM)* dengan 3 (tiga) klasifikasi: positif, negatif dan netral
4. Pada penelitian ini akan membandingkan ekstraksi fitur *count vector*, *tf-idf word level*, *tf-idf ngram level* dan *tf-idf char level*.
5. Tahap penelitian hanya sampai pada perbandingan nilai akurasi tertinggi dari opini pengguna pada media sosial *facebook* dan *instagram*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

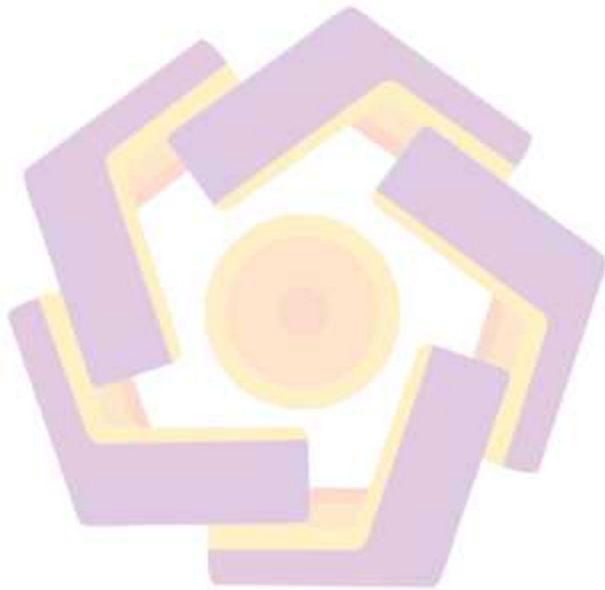
1. Untuk mengetahui metode alternatif lain dalam scraping data pada media sosial *facebook* dan *instagram*.
2. Untuk mengetahui pada media sosial manakah algoritma *Support Vector Machine* menghasilkan nilai akurasi tertinggi.
3. Untuk mengetahui fitur apa saja yang mempengaruhi nilai akurasi tertinggi pada metode *Support Vector Machine* pada *facebook* dan *instagram*.

1.5 Manfaat Penelitian

Manfaat akhir dari penelitian ini nantinya yang diharapkan oleh penulis adalah sebagai berikut :

1. Secara teoritis, memperkaya penelitian di bidang data mining dan text mining di bidang media sosial
2. Secara praktis adalah sebagai berikut:
 - a. Dapat mengetahui teknik alternatif lain untuk scraping data pada *facebook* dan *instagram* karena beberapa media scraping data di *instagram* dan *facebook* sudah banyak yang ditutup aksesnya untuk keamanan.
 - b. Dapat mengetahui pada media sosial manakah algoritma *Support Vector Machine* menghasilkan nilai akurasi tertinggi.
 - c. Dapat mengetahui fitur apa saja yang mempengaruhi nilai akurasi tertinggi pada metode *Support Vector Machine* pada *facebook* dan *instagram*.

- d. Mendapatkan informasi bagi pelaku industri dan pihak terkait di media sosial mana di Kota Kediri yang tepat sebagai media penyebaran informasi dan promosi sehingga lebih fokus pada media sosial tersebut



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian sebelumnya sangatlah penting bagi penulis sebagai bahan kajian untuk mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan penulis lakukan serta untuk menghindari terjadinya duplikasi yang dilakukan penulis. Tinjauan pustaka juga bertujuan untuk menunjukkan bahwa penelitian yang penulis lakukan sangatlah bermanfaat dan memiliki arti penting sebagai kontribusi penelitian terhadap ilmu pengetahuan. Berikut ini adalah beberapa ulasan jurnal tentang penelitian terdahulu berkenaan dengan data dan metode yang digunakan penulis sebagai acuan.

Dalam penelitian yang dilakukan oleh Akbar (2018) tentang penerapan metode Support Vector Machine (SVM) dalam media sosial twitter dan facebook AMIKOM Yogyakarta. Hasil eksperimen menunjukkan facebook memiliki presentase lebih tinggi daripada twitter.

Abu (2018) dalam penelitiannya tentang analisis sentiment komentar facebook dengan K-Nearest Neighbor (KNN) pada jasa ekspedisi barang J&T Express diperoleh hasil akurasi tertinggi 79,21 dan akurasi terendah 70,3 dengan jumlah dokumen training 101.

Wanda (2018) dalam penelitiannya tentang analisis sentiment cyberbullying pada komentar instagram dengan menggunakan Support Vector Machine (SVM) dengan tingkat akurasi 90%.

Eko (2019) dalam penelitiannya tentang analisis sentiment calon presiden Indonesia 2019 berdasarkan komentar public di facebook dengan menggunakan Naïve Bayes dengan hasil Joko Widodo 4242 komentar positif dan 758 komentar negatif. Sedangkan Prabowo Subianto memperoleh 3824 komentar positif dan 1176 komentar negatif.

Moehammad (2020) dalam penelitiannya tentang analisis sentiment tweet di Arab Saudi tentang pencegahan Pemerintah terhadap Covid-19. Moehammad menggunakan Naïve Bayes dalam penelitiannya dan dihasilkan bahwa kicauan di tweeter yang diteliti belum mewakili populasi masyarakat Arab Saudi karena banyak kicauan yang mengesampingkan tanda pagar sehingga tidak masuk dalam mesin analisis yang digunakan dan diperlukan survey di media sosial lain.

James (2020) dalam penelitiannya yang berjudul analisis sentiment di media sosial dengan metode maximum entropy dihasilkan bahwa diperlukan kontribusi manusia dalam maximum entropy sehingga nilai akurasi tinggi.

2.2. Keaslian Penelitian

Tabel 2.1
Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada *Facebook* dan *Instagram*
Menggunakan *Support Vector Machine*

No	Judul	Peneliti, Media Publikasi, Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1.	Analisis sentimen media sosial Univ.Amikom Yogyakarta sebagai sarana penyebaran informasi menggunakan Algoritma Klasifikasi SVM	Akbar Maulana dkk, Seminar Nasional Teknologi Informasi dan Multimedia, 2018	Untuk mengetahui prosentase media sosial amikom Yogyakarta	Facebook memiliki prosentase lebih tinggi daripada twiter	Hanya difokuskan pada media social facebook dan twiter	Penulis menggunakan media sosial facebook dan instagram dengan SVM
2.	Analisis sentimen data komentar sosial media fb dengan K-Nearest Neighbor (Studi kasus pada akun jasa ekspedisi barang J&T Ekspres Indonesia)	Abu Salam, dkk, Prosiding SINTAK 2018	Mengklasifikasi komentar-komentar pelanggan yang masuk kedalam facebook page J&T Ekspres Indonesia masuk kedalam Sentiment negatif atau Sentiment positif dengan metode KNN	Hasil akurasi tertinggi dengan metode KNN 79,21 dan akurasi terendah 70,3 dengan jumlah dokumen training 101	Menyarankan algoritma lain untuk membandingkan	Penulis Menggunakan algoritma SVM yang akurasinya lebih tinggi daripada KNN

Tabel 2.1

Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada *Facebook* dan *Instagram*
Menggunakan *Support Vector Machine* (Lanjutan)

3.	Analisis sentimen cyberbullying pada komentar instagram dengan metode klasifikasi Support Vector Machine	Wanda Athira dkk, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2018	Menganalisis komentar cyberbullying pada instagram menggunakan SVM	Didapatkan tingkat akurasi 90% dengan mengklasifikasi menjadi komentar positif dan negatif	Hanya menampilkan 2 klasifikasi (positif dan negatif)	Menggunakan 3 klasifikasi (positif, negative dan netral)
4.	Analisis sentimen calon presiden Indonesia 2019 berdasarkan komentar Publik di Fb	Eko Budi Santoso, dkk, Jurnal Eksplora Informatika, 2019	Menganalisis komentar pada facebook tentang calon presiden 2019 dengan Naïve Bayes	Didapatkan hasil Joko Widodo 4242 komentar positif dan 758 komentar negatif. Sedangkan Prabowo Subianto memperoleh 3824 komentar positif dan 1176 komentar negatif.	Hanya pada media sosial facebook	Pada media social facebook dan instagram

Tabel 2.1

Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada *Facebook* dan *Instagram*
Menggunakan *Support Vector Machine* (Lanjutan)

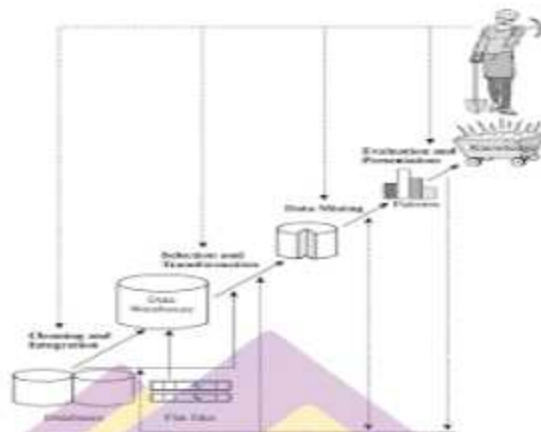
5.	Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19	Moehammad Al Hajji, Preprints, 2020	Menganalisis tweet tentang pencegahan Pemerintah terhadap Covid-19 di Arab Saudi menggunakan Naïve Bayes	Kicauan di tweeter yang diteliti belum mewakili populasi masyarakat Arab Saudi karena banyak kicauan yang mengesampingkan tanda pagar sehingga tidak masuk dalam mesin analisis yang digunakan dan diperlukan survey di media sosial lain.	Hanya pada social media twiter	Menggunakan 2 media (instagram dan facebook)
6.	Studying social media sentiment using human validated analysis	James Lappeman, Elsevier, 2020	analisis sentiment di media sosial dengan metode maximum entropy	diperlukan kontribusi manusia dalam maximum entropy sehingga nilai tinggi	Difokuskan pada social media twiter	Dengan menggunakan 2 sosial media (instagram dan facebook)

2.3. Landasan Teori

2.3.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan machine learning untuk mengekstraksi serta mengidentifikasi informasi yang bermanfaat untuk pengetahuan yang terkait dari berbagai database besar (Turban et al, 2005). Menurut Tan et al (2006) data mining adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *knowledge discovery*.

Data mining adalah bagian integral dari penemuanpengetahuan dalam database (KDD) yang merupakan proses keseluruhan mengubah data mentah menjadi pola-pola data menarik yang merupakan informasi yang dibutuhkan oleh pengguna sebagai pengetahuan. Untuk mengetahui proses *knowledge discovery* dalam database bisa dilihat pada Gambar 2 berikut



Gambar 2.1 Tahapan Proses KDD (Han dan Kamber 2006)

Han dan Kamber (2006) menyatakan bahwa KDD terdiri dari langkah-langkah sebagai berikut:

1. Data cleaning adalah proses menghapus data yang tidak konsisten dan menghilangkan noise.
2. Data integration adalah proses menggabungkan data apabila memiliki sumber data dalam sistem data mining tersebut.
3. Data selection adalah pengambilan data yang relevan yang akan digunakan dalam proses data mining.
4. Data transformation adalah proses dimana data ditransformasikan menjadi bentuk-bentuk yang sesuai untuk proses dalam data mining.
5. Data mining adalah suatu proses yang penting dengan melibatkan metode-metode untuk menghasilkan suatu pola data.
6. Pattern evaluation adalah proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada didalam data itu sendiri.

7. Knowledge representation adalah proses visualisasi dan teknik menyajikan knowledge digunakan untuk menampilkan knowledge hasil mining kepada pengguna.

2.3.2 Machine Learning

Machine Learning adalah bidang yang mempelajari pengembangan algoritma komputer untuk mengubah data menjadi aksi yang cerdas atau secara singkat dapat juga diartikan sebagai proses mengubah data menjadi informasi (Suyanto, 2017).

Menurut Mohri et al (2012) machine learning dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajar. Dalam pembelajaran machine learning, terdapat beberapa skenario-skenario. Seperti:

1. Supervised Learning

Penggunaan skenario supervised learning, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

2. Unsupervised Learning

Penggunaan skenario Unsupervised Learning, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokan data berdasarkan karakteristik-karakteristik yang ditemui.

3. Reinforcement Learning

Pada skenario reinforcement learning fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

2.3.3 Text Mining

Feldman dan Sanger (2007) menyatakan text mining adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis. Text mining atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan data mining dimana data mining mengolah data yang sifatnya terstruktur (Jamil, 2017).

Perbedaan antara text mining dengan data mining terletak pada sumber data yang digunakan. Dalam text mining pola-pola yang diekstrak dari data tekstual yang tidak terstruktur bukan berasal dari suatu database. Beberapa kesamaannya adalah data yang digunakan merupakan data besar dan data berdimensi tinggi dengan struktur yang terus berubah. Dalam data mining data yang diolah adalah data yang terstruktur dari proses *warehousing* sehingga lebih mudah diproses oleh mesin/komputer.

Analisis teks lebih sulit karena teks biasanya hanya digunakan sebagai konsumsi manusia secara langsung bukan digunakan untuk mesin/komputer.

Ditambah struktur teks yang kompleks, struktur yang tidak lengkap, bahasa yang berbeda, dan arti yang tidak standar. Oleh sebab itu pada umumnya digunakan Natural Language Processing untuk analisis teks yang tidak berstruktur tersebut. Tahapan-tahapan dalam text mining secara umum adalah text preprocessing dan feature selection (Feldman & Sanger, 2007).

Dalam melakukan text mining, teks dokumen yang digunakan harus dipersiapkan terlebih dahulu, setelah itu baru dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau dataset mentah disebut juga dengan proses text preprocessing. Text preprocessing berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Secara umum proses yang dilakukan dalam tahapan preprocessing adalah sebagai berikut:

a. *Case Folding*

Case folding adalah proses penyamaan *case* dalam sebuah dokumen. Dalam hal ini dalah menyamakan karakter pada data. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (dalam hal ini huruf kecil atau lowercase).

b. *Tokenizing*

Tokenizing adalah proses memecah kalimat menjadi kata-kata yang dilakukan untuk menjadikan sebuah kalimat menjadi lebih bermakna. Tahap pertama yang dilakukan adalah normalisasi kata dengan mengubah semua karakter huruf menjadi huruf kecil atau toLowerCase. Proses tokenisasi diawali

dengan menghilangkan delimiter-delimiter yaitu simbol dan tanda baca yang ada pada teks tersebut seperti @, \$, &, tanda titik (.), koma (,) tanda tanya (?), tanda seru (!). Tahap tokenisasi selanjutnya yaitu proses penguraian teks yang semula berupa kalimat-kalimat yang berisi kata-kata.

Proses pemotongan string berdasarkan tiap kata yang menyusunnya, umumnya setiap kata akan terpisahkan dengan karakter spasi, proses tokenisasi mengandalkan karakter spasi pada dokumen teks untuk melakukan pemisahan. Hasil dari proses ini adalah kumpulan kata saja (Putri, 2016).

c. *Stopword Removal*

Stopword Removal adalah tahap mengambil kata-kata penting dari hasil token. Algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata yang penting) dapat digunakan pada tahap ini. Stopword adalah kata-kata yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuang. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya (Putri, 2016). Dalam filtrasi ini menggunakan stoplist/stopword agar kata-kata yang kurang penting dan sering muncul dalam suatu dokumen dibuang sehingga hanya menyisakan kata-kata yang penting dan mempunyai arti yang diproses ke tahap selanjutnya.

d. *Stemming*

Stemming adalah Proses pengembalian suatu kata berimbuhan ke dalam bentuk kata dasarnya (*Root*). Sama halnya dengan *stopword* kinerja *stemming* bervariasi dan sering bergantung pada dokumen bahasa yang digunakan.

2.3.4 Analisis Sentimen

Analisis sentiment atau bisa juga disebut opinion mining merupakan sebuah cabang penelitian di domain text mining yang mulai banyak dilakukan pada tahun 2013. Lee dan Pang menjelaskan analisis sentiment atau dikenal sebagai opinion mining adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi (Lee & Pang, 2008).

Secara umum, opinion mining diperlukan untuk mengetahui sikap seorang pembicara atau penulis sehubungan dengan beberapa topik atau polaritas kontekstual keseluruhan dokumen. Sikap yang diambil mungkin menjadi pendapat atau penilaian atau evaluasi (teori appraisal), keadaan afektif (keadaan emosional penulis saat menulis) atau komunikasi emosional (efek emosional penulis yang ingin disampaikan pada pembaca) (Saraswati, 2011).

Analisis sentimen dapat digunakan dalam berbagai kemungkinan domain, dari produk konsumen, jasa kesehatan, jasa keuangan, peristiwa sosial dan politik pada pemilu. Kecendrungan penelitian tentang analisis sentimen berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negatif. Pendapat mewakili hampir semua aktivitas manusia, karena pendapat dapat mempengaruhi terhadap perilaku seseorang. Setiap kali kita perlu membuat keputusan, kita ingin tahu pendapat orang lain. Dalam dunia nyata, bisnis dan organisasi selalu ingin melihat opini publik tentang suatu produk atau jasa. (Liu, 2012).

2.3.5 Media Sosial

Pada dasarnya, media sosial adalah situs atau layanan daring (online) yang memungkinkan penggunanya tak hanya mengonsumsi, tapi juga berpartisipasi membuat, mengomentari, dan menyebarkan beragam konten dalam berbagai format: teks, gambar, audio, atau video. Dengan media sosial, penggunanya bisa membangun percakapan, bahkan komunitas, karena media sosial juga mempermudah pertemuan beberapa atau banyak orang dengan minat sama. Media ini juga memudahkan pengelola usaha, organisasi masyarakat, sampai lembaga pemerintah untuk terkoneksi langsung dengan publik. Media sosial yang populer diantaranya adalah Facebook, Twitter, Instagram, YouTube, atau Pinterest. Berikut ini 6 (enam) jenis media sosial:

1. Layanan blog

Blog secara ringkas bisa dipahami sebagai jurnal pribadi di internet, untuk berbagi catatan atau pandangan penggunanya tentang beragam hal. Penggunaannya lazim disebut sebagai narablog (blogger). Contoh: WordPress, Blogger. Untuk pemasaran, blog bisa digunakan dalam kegiatan Content Marketing, seperti business blogging, mengedukasi sasaran pasar tentang topik tertentu atau keunggulan produk atau jasa yang ditawarkan, atau memperkuat SEO atau relasi brand dengan topik tertentu.

2. Layanan jejaring sosial (social network)

Jenis layanan yang fokus pada terbangunnya jejaring di antara penggunanya untuk saling berbagi pesan, informasi, foto, atau video. Model relasi antar pengguna

yang lumrah berbentuk pertemanan dengan cara saling Add atau Connect. Contoh: Facebook, LinkedIn.

3. Layanan blog mikro (microblogging)

Meski kegunaannya serupa, tapi jenis media ini lebih ringkas, hingga memengaruhi alur interaksinya yang jadi lebih cepat dibandingkan blog. Contoh: Twitter. Microblogging bisa dioptimalkan sebagai channel untuk berinteraksi cepat dengan audiens yang disasar, hingga berbagi informasi ringkas yang penting diketahui konsumen atau pengguna lain di channel terkait. Contohnya, jadi saluran untuk menangani keluhan atau pertanyaan konsumen atau pelanggan.

4. Layanan berbagi media (media sharing)

Kalau kita suka menonton YouTube atau mendengar Soundcloud, keduanya tergolong dalam jenis ini; yang fokus utamanya memang untuk berbagi konten media seperti foto, audio, atau video. Contoh lain: Instagram, Flickr. Pemasar bisa menggunakan jenis ini untuk berbagi konten-konten yang kuat secara visual; seperti foto produk, kegiatan brand, sampai konten seperti infografik atau video.

5. Layanan forum

Bisa dibilang sebagai jenis media sosial klasik yang sudah dikenal sejak lama. Layanan ini jadi tempat pengguna bisa memperbincangkan hal atau topik spesifik dengan pengguna lain di dalam ruang diskusi. Contoh: Kaskus, Quora.

6. Layanan kolaborasi

Seperti namanya, layanan ini memberi kesempatan penggunanya untuk berkolaborasi dalam memuat, menyunting, atau mengoreksi konten. Contoh: Wikipedia.

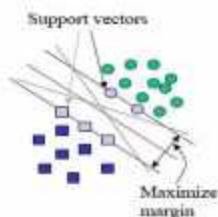
2.3.6 Covid-19

Covid-19 merupakan penyakit yang disebabkan oleh Coronavirus. Sedangkan coronavirus sendiri merupakan keluarga besar virus yang menyebabkan penyakit pada manusia dan hewan. Pada manusia biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti *Middle East Respiratory Syndrome (MERS)* dan Sindrom Pernafasan Akut Berat/*Severe Acute Respiratory Syndrome (SARS)*. Coronavirus jenis baru yang ditemukan pada manusia sejak kejadian luar biasa muncul di Wuhan Cina, pada Desember 2019, kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV2)*, dan menyebabkan penyakit *Coronavirus Disease-2019 (COVID-19)*.

(<http://infocovid19.jatimprov.go.id/> diakses 12 April 2020)

2.3.7 Support Vector Machine (SVM)

Support Vector Machines (SVM) adalah seperangkat metode pembelajaran yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi, tidak hanya itu teknik ini dapat melakukan prediksi dan penilaian tentang sebuah sistem.



Gambar 2.2 Support Vector Machine

SVM sangat terkenal dengan komunitas klasifikasi pola yang secara default, para periset di bidang ini menggunakannya sebagai penggolongan dasar untuk menetapkan superioritas *classifier* yang diusulkan oleh mereka (Murty, 2016). Tujuan dari SVM adalah memberikan nilai dari banyaknya kemunculan sebuah kata atau dapat mengklasifikasi komentar positif dan negatif. Pada teknik *Support Vector Machine* (SVM) akan dikelompokkan 2 data atau lebih dan akan dibuat garis untuk memisahkan kelompok data tersebut. Garis ini dikenal dengan *hyperplane*. Teknik SVM bertujuan untuk mencari hyperplane yang optimal. Formula untuk algoritma SVM adalah seperti di bawah ini dimana x_i merupakan data masukan dan y_i merupakan keluaran, sedangkan w dan b merupakan parameter yang kita cari nilainya.

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 \\ & \text{Subject to} \\ & y_i(wx_i + b) \geq 1, \quad i = 1, \dots, \ell, \end{aligned}$$

Gambar 2.3 Formula untuk *Support Vector Machine*

2.3.8 Python

Python adalah bahasa pemrograman tujuan umum yang ditafsirkan tingkat tinggi. Dibuat oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991, filosofi desain Python menekankan keterbacaan kode dengan penggunaan spasi putih yang signifikan. Python adalah bahasa pemrograman multi-paradigma. Python mendukung pemrograman fungsional dan pemrograman berorientasi aspek

termasuk dengan metaprogramming dan metaobjects. Banyak paradigma lain yang didukung melalui ekstensi, termasuk desain berdasarkan kontrak dan pemrograman logika. Desain Python menawarkan beberapa dukungan untuk pemrograman fungsional dalam tradisi Lisp. Memiliki fungsi filter, map, dan reduce;daftar pemahaman, kamus, set, dan ekspresi generator.

2.3.9 WebHarvy

Webharvy adalah Salah satu teknik pengumpul data pada web atau media sosial yang berbentuk perangkat lunak cerdas yang secara cerdas dapat mengikis data dari halaman web atau media sosial dan menyimpan konten yang diekstrak dalam format yang berbeda (.csv).



Gambar 2.4 Tampilan WebHarvy

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

3.1.1 Jenis

Jenis penelitian yang digunakan adalah penelitian eksperimen dan studi kasus. Penelitian eksperimen adalah penelitian yang dilakukan dengan menerapkan serangkaian tindakan untuk membuktikan suatu konsep. (Sugiyono, 2015). Dalam hal ini penulis melakukan eksperimen melalui proses *scraping* data *facebook* dan *instagram* kemudian mengklasifikasikannya menggunakan *Support Vector Machine*. Studi kasus dilaksanakan pada media sosial *facebook* dan *instagram* akun AG243.

3.1.2 Sifat Penelitian

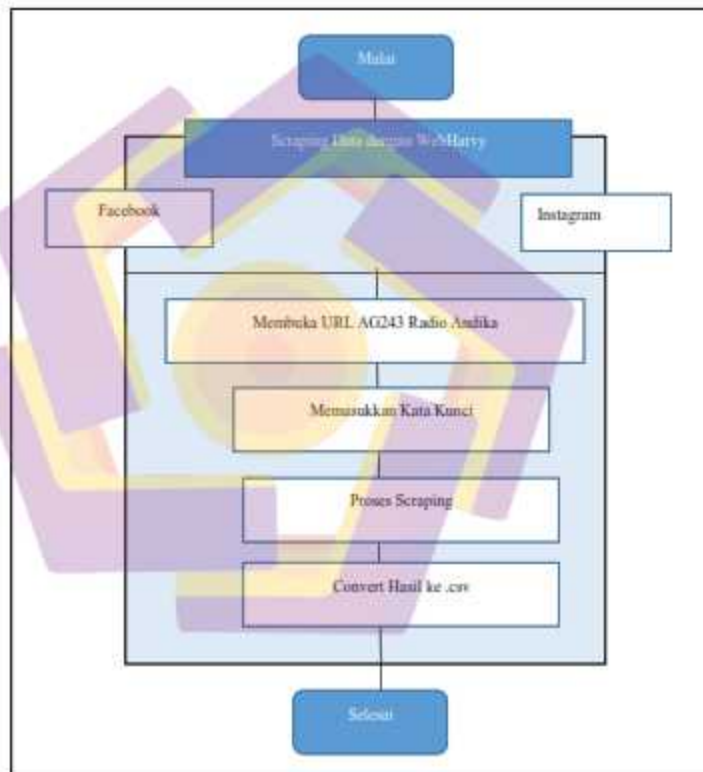
Penelitian yang dilakukan penulis bersifat deskriptif, dimana data dijelaskan dalam deskriptif angka dan tabel atau diagram.

3.1.3 Pendekatan Penelitian

Setelah mengolah data, penulis menganalisisnya dengan pendekatan kuantitatif yang dijelaskan dengan hasil penghitungan angka dan tabel atau diagram.

3.2 Metode Pengumpulan Data

Data dikumpulkan dengan cara *scraping* data komentar dengan menggunakan *webharvy* yang ada pada *facebook* dan *instagram* AG243 yang berisi informasi covid-19 pada 18 Juli s/d 18 Agustus 2020. Sedangkan alur pengumpulan data adalah seperti di bawah ini:



Gambar 3.1 Alur Pengumpulan Data

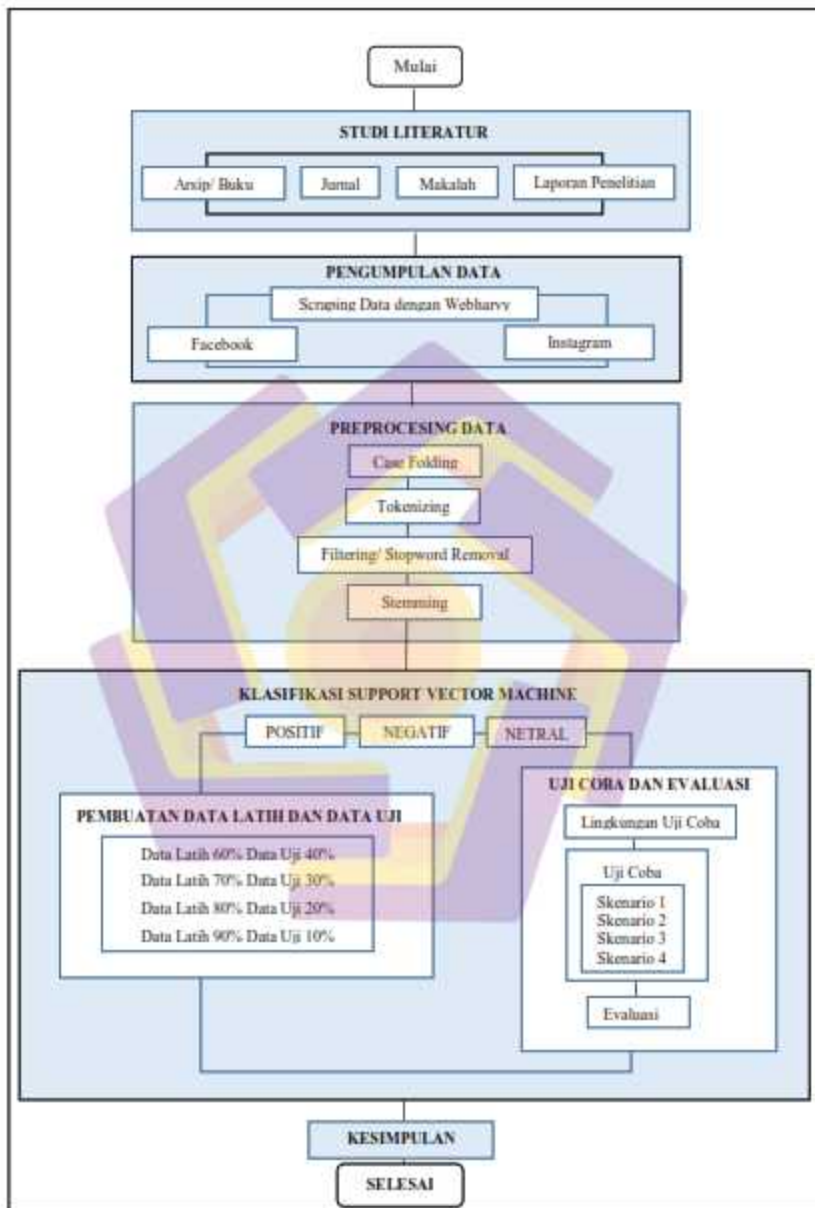
3.3 Metode Analisis Data

Dalam penelitian ini, peneliti menggunakan webharvy, python dan Microsoft Excel. Adapun metode analisis data yang digunakan dalam penelitian ini yaitu:

1. Analisis deskriptif untuk memberikan gambaran umum komentar di *facebook* dan *instagram* terhadap penyebaran informasi Covid-19 oleh akun AG243.
2. Analisis deskriptif di media sosial manakah AG243 tepat dalam penyebaran informasi khususnya tentang Covid-19 dan informasi lain pada umumnya
3. Analisis Sentimen dengan *metode machine learning* yaitu *Support Vector Machine (SVM)* yang digunakan untuk mengklasifikasikan komentar yang berbentuk positif, negatif dan netral.

3.4 Alur Penelitian

Berdasarkan uraian sebelumnya, dapat ditentukan alur penelitian sebagai berikut:



Gambar 3.2 Alur Penelitian

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

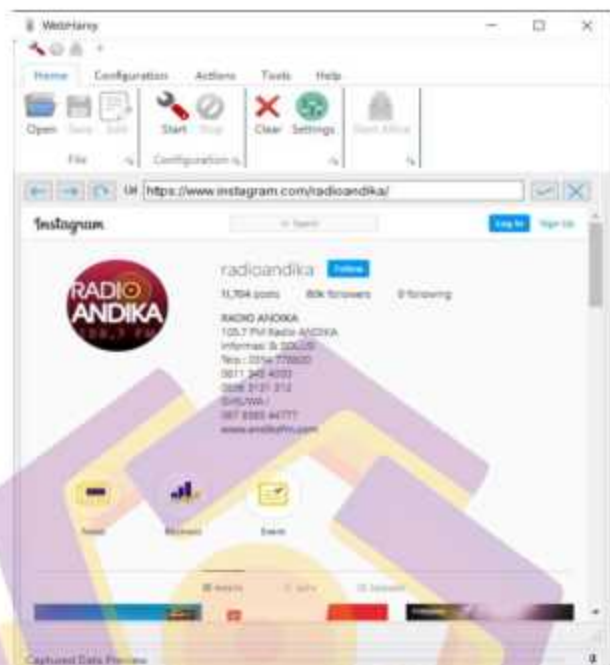
4.1 Tahap Scraping Data Komentar Facebook dan Instagram

Pengambilan data pada *facebook* dan *instagram* menggunakan tools *web harvy*. *Webharvy* merupakan salah satu aplikasi web *scraping* yang bisa digunakan untuk mengambil data komentar pada web dan media sosial. Data diambil dari tanggal 18 Juli 2020 s/d 18 Agustus 2020.



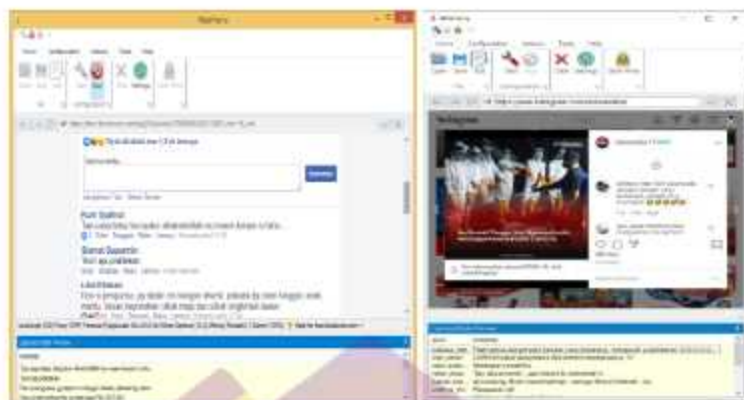
Gambar 4.1 Tampilan *Facebook* pada *webharvy*

Pada gambar 4.1 lembar *webharvy*, kita tinggal memasukkan url yang akan kita *scraping* datanya. Disini penulis memasukkan url facebook AG243, setelah muncul kita bisa memasukkan kata kunci, disini penulis memasukkan kata kunci "covid19, covid-19, corona, covid dan pandemi" secara bergantian. Ketika kata kunci sudah dimasukkan tinggal klik start. Maka *webharvy* akan melakukan proses *scraping*.



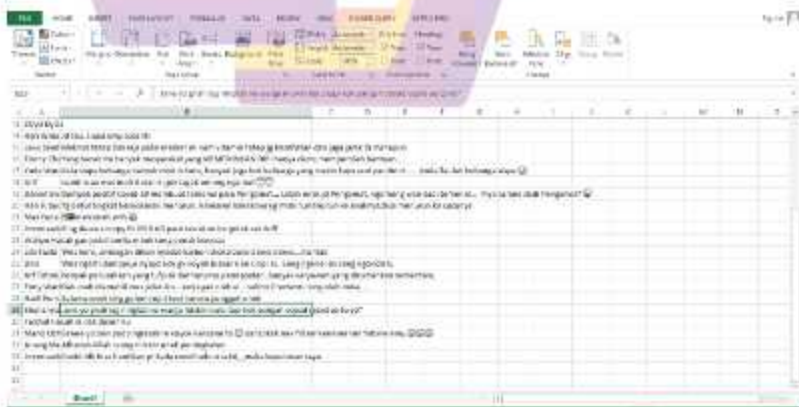
Gambar 4.2 Tampilan *Instagram* pada *Webharvy*

Sama halnya pengambilan data pada *facebook*, pada gambar 4.2 pengambilan data pada *instagram* radio andika di aplikasi *webharvy* juga dilakukan dengan memasukkan url *instagram* radio andika. Kemudian melakukan searching sesuai kata kunci yang kita gunakan serta proses yang sama dengan pengambilan data pada *facebook*.



Gambar 4.3 Proses *Scraping* data FB dan IG pada *webharvy*

Setelah melakukan proses *scraping* data seperti pada gambar 4.3 maka akan muncul hasil *scraping*-nya pada kolom captured data. Setelah selesai yaitu ditandai dengan berhentinya proses loading *scraping*, kita tinggal klik stop. Dan bisa mengulangi untuk kata kunci yang lain. Setelah itu kita bisa melakukan proses convert ke.csv yang akan bisa kita buka pada file microsoft excell.



Gambar 4.4 Hasil *Scraping* Data

Setelah data di convert ke lembar .csv seperti halnya tertera pada gambar 4.4 maka dapat diketahui bahwa hasil *scraping* data pada media *facebook* dan *instagram* adalah seperti yang tertera pada Tabel 4.1.

Tabel 4.1 Hasil Scraping Data Facebook dan Instagram

Kata Kunci	Facebook	Instagram
covid-19	267	87
covid19	361	126
covid	671	435
corona/ korona	171	274
pandemi	489	653
Jumlah	1959	1575

Pada tabel 4.1 dapat diketahui bahwa pada kata kunci covid-19 didaatkan hasil *scraping* data 267 komentar pada *facebook* dan 87 pada *instagram*. Sedangkan dengan kata kunci covid19 terdapat 361 komentar *facebook* dan 126 komentar *instagram*. Dengan kata kunci covid terdapat 671 komentar *facebook* dan 435 komentar *instagram*. Dengan kata kunci corona/ korona 171 komentar *facebook* dan 274 komentar *instagram*. Serta dengan kata kunci pandemi terdapat 489 komentar *facebook* dan 653 komentar *instagram*. Total semua komentar adalah 1959 komentar pada *facebook* dan 1575 komentar pada *instagram*.

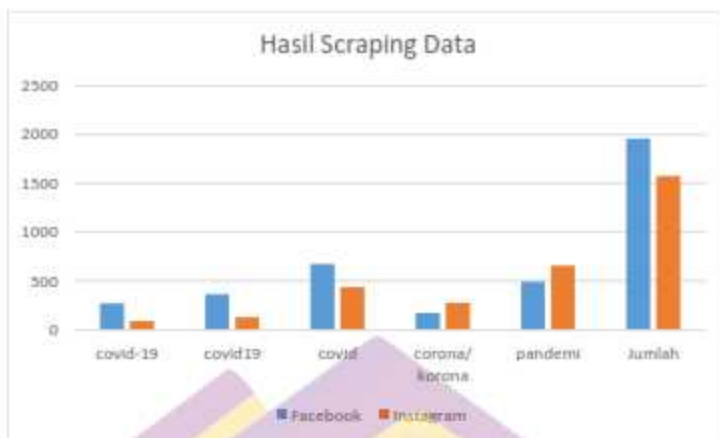


Diagram 4.1 hasil Scraping Data Facebook dan Instagram

Sebagaimana diagram 4.1 hasil scraping data facebook dan instagram dapat diketahui dengan masa rentang waktu yang sama yaitu pada 18 Juli 2020 s/d 18 Agustus 2020, facebook memiliki data komentar lebih besar daripada facebook

Setelah selesai convert ke.csv hasil *scraping* data facebook dan instagram dengan *webharvy* bisa kita buka di microsoft excell dan bisa kita lakukan tahap *preprocessing data*.

4.2 Pre Processing Facebook dan Instagram

Pada penelitian ini menerapkan *text pre processing* dalam melakukan pengolahan data yang akan digunakan dalam proses analisis sentiment. Data yang diperoleh dari opini pengguna *Facebook* dan *Instagram* memiliki bentuk yang tidak terstruktur karena masih banyak mengandung kata-kata yang tidak ada maknanya sehingga sulit untuk dilakukan analisis secara

langsung. Sehingga hal ini perlu dilakukan seleksi dalam mempermudah analisis data dan data yang digunakan lebih terstruktur serta memiliki keseragaman untuk proses mining lebih lanjut. Penulis menggunakan script Python untuk mempermudah dalam melakukan proses *preprocessing*.

Tabel 4.2 Contoh komentar pada Facebook dan Instagram AG243

NO	User	Komentar
1	Lcxxxxxx	Maknya tetep bekerja pake masker minum vitamin tetep jg kesehatan dan jaga jarak di manapun!?!
2	Tonxxxxx	Yang bener itu banyak masyarakat yang MEMISKINKAN DIRI dicovid ini hanya demi memperoleh bantuan...wkwkwkw
3	Yudxxxxx	Kata siapa keluarga banyak miskin baru, banyak juga kok keluarga yang makin kaya saat pandemi Anda Taulah keluarga siapa ☹.

Pada tahap *text preprocessing*, metode text mining akan dilakukan untuk pembersihan data diantaranya adalah *case folding*, *tokenizing*, *filtering* atau *stopword removal* dan *stemming*.

4.2.1 Case Folding

Tahap *preprocessing* yang pertama adalah *case folding*. *Case folding* adalah proses mengubah semua karakter dalam dokumen menjadi tema yang sama. Contohnya, dalam dokumen teks tidak semua penggunaan huruf kapital dilakukan dengan konsisten. Disini penulis menggunakan script python di bawah ini seperti pada gambar 4.2.

```
train_x = train_csv['0'].str.lower()
train_y = train_csv['1'].str.lower()
test_x = test_csv['0'].str.lower()
test_y = test_csv['1'].str.lower()
```

Gambar 4.5 Script Python untuk *case folding*

Peran case folding dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil). Dalam case folding, hanya karakter huruf a sampai z yang diterima, sedangkan karakter lain tidak diterima dan dianggap pembatas. Contoh penggunaan case folding dapat dilihat pada tabel 4.2 berikut:

Tabel 4.2 Proses Case Folding

NO	Komentar Awal	Komentar Hasil Case Folding
1	Maknya tetep bekerja pake masker minum vitamin tetep jg kesehatan dan jaga jarak di manapun?!	maknya tetep bekerja pake masker minum vitamin tetep jg kesehatan dan jaga jarak di manapun
2	Yang bener itu banyak masyarakat yang MEMISKINKAN DIRI dicovid ini hanya demi memperoleh bantuan....wkwkwkw	yang bener itu banyak masyarakat yang memiskinkan diri dicovid ini hanya demi memperoleh bantuan wkwkwkw
3	Kata siapa keluarga banyak miskin baru, banyak juga kok keluarga yang makin kaya saat pandemi Anda Taulah keluarga siapa 😊.	kata siapa keluarga banyak miskin baru banyak juga kok keluarga yang makin kaya saat pandemic anda taulah keluarga siapa

4.2.2 Tokenizing

Tahap selanjutnya yaitu *tokenizing* atau bisa juga disebut *parsing*. *Tokenizing* adalah proses pemotongan dokumen menjadi bagian-bagian kata yang disebut token. Spasi digunakan untuk memisahkan antar kata tersebut. Sedangkan kata-kata yang tidak dibutuhkan akan dihilangkan melalui proses filtering dari hasil tokenizing.

```
for k in range(len(train_X)):
    train_X[k] = nltk.tokenize.word_tokenize(train_X[k])

for k in range(len(test_X)):
    test_X[k] = nltk.tokenize.word_tokenize(test_X[k])
```

Gambar 4.6 Script tokenizing

Dengan hasil tokenizing seperti pada gambar di bawah ini

makanya	tetap
bekerja	pake
masker	minum
vitamin	tetep
lg	kesehatan
dan	jaga
jarak	di
manapun.	

Gambar 4.7 Contoh hasil *Tokenizing*

4.2.3 *Stopword Removal atau Filtering*

Pada proses ini menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist / stopwords adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Kata-kata yang dihilangkan yaitu: - penghubung antar kata, seperti: dan, atau, serta - preposisi, seperti: di, ke, pada - kata-kata yang tidak diinginkan. Dalam hal ini script Python yang digunakan adalah seperti pada gambar 4.8

```
factoryStopRemove = StopWordRemoverFactory()
stopword = factoryStopRemove.create_stop_word_remover()
= for z in range(len(train_X)):
    train_X[z] = stopword.remove(train_X[z])

= for z in range(len(test_X)):
    test_X[z] = stopword.remove(test_X[z])
```

Gambar 4.8 Script filtering atau stopwords removal pada Python

Dengan contoh hasil stopword removal seperti pada gambar 4.9



Gambar 4.9 Proses Filtering atau Stopword Removal

4.2.4 Stemming

Setelah proses *tokenizing* dan *filtering* selanjutnya adalah proses *stemming* yang digunakan untuk mencari root dari kata hasil proses *filtering*. Contoh proses *stemming* dapat dilihat pada gambar 4.10 di bawah ini.



Gambar 4.10 Proses Stemming

Dalam proses ini ada pengembalian kata berimbuhan ke kata dasarnya (root). Dalam contoh di atas yang kembali ke kata dasarnya ada kata “bekerja” menjadi “kerja” dan kata “kesehatan” menjadi “sehat” dengan script python seperti pada gambar 4.11.

```
factoryStemm = StemmerFactory()
stemmer = factoryStemm.create_stemmer()
* for y in range(len(train_X)):
    train_X[y] = stemmer.stem(train_X[y])
* for y in range(len(test_X)):
    test_X[y] = stemmer.stem(test_X[y])
```

Gambar 4.11 Script *stemming* pada Python.

4.3 Klasifikasi Support Vector Machine

4.3.1 Pembuatan Data Latih

Pembuatan data latih diperlukan untuk melatih kinerja *machine learning*. Data latih akan mempengaruhi tingkat akurasi yang dihasilkan oleh machine learning. Pada fase latih (learning), sebagian data yang telah diketahui kelas datanya (training set) digunakan untuk membentuk model. Data latih dibuat dengan cara mengklasifikasikan secara manual seluruh data yang didapatkan dari komentar facebook dan instagam ke dalam komentar positif, negatif maupun netral . Pembuatan data latih dilakukan dengan Selain itu peneliti juga menentukan opsi lain dengan menyeimbangkan jumlah data latih dengan data uji.

Tabel 4.3 Pembagian jumlah Data Latih dan Data Uji

Opsi	Jumlah Dataset		Jumlah Data Latih/ Training			Jumlah Data Uji/ Testing		
	Face book	Insta gram	Prose ntase	Face book	IG	Prosent ase	Facebook	IG
Skenario 1	1027	1000	60%	616	600	40%	411	400
Skenario 2			70%	719	700	30%	308	300
Skenario 3			80%	822	800	20%	205	200
Skenario 4			90%	924	900	10%	103	100

Contoh klasifikasi data latih yang dilakukan secara manual dapat dilihat pada tabel 4.3. Dalam penelitian ini peneliti mengklasifikasikan komentar ke dalam kelas positif, kelas negatif dan netral. Komentar yang memuat opini positif mengenai informasi covid-19 akan diklasifikasikan ke dalam kelas positif. Sedangkan untuk kalimat yang memuat opini negatif dan keluhan akan diklasifikasikan ke dalam kelas negatif. Serta yang memuat opini netral akan dimasukkan kedalam kelas netral.

Tabel 4.4 Contoh Evaluasi Data latih Secara Manual

Klasifikasi	Komentar
Positif	tetep bekerja pake masker minum vitamin tetep jg kesehatan dan jaga jarak di manapun!?!
Negatif	kata siapa keluarga banyak miskin baru, banyak juga kok keluarga yang makin kaya saat pandemi anda taulah keluarga siapa ☹️.
Netral	Aq tak no comment saja

4.3.2 Uji coba dan Evaluasi Klasifikasi *Support Vector Machine*

1. Lingkungan Uji Coba

Lingkungan uji coba pada penelitian ini berupa perangkat keras serta perangkat lunak yang digunakan untuk melakukan uji coba text mining pada media

sosial facebook dan instagram AG243 Radio Andika. Lingkungan uji coba ini berupa komputer dengan spesifikasi seperti pada tabel 4.5 di bawah ini.

Tabel 4.5 Lingkungan Uji Coba

Perangkat	Spesifikasi
Perangkat Keras	Processor Intel® Celeron® 2957U (@ 1.40GHz 1.40GHz Memory 2GB
Perangkat Lunak	Sistem Operasi : Microsoft Windows 8.1 Pro Perangkat Pengembang : Anaconda Perangkat Pembantu : Spyder, Microsoft Excell 2013

2. Uji Coba dan Evaluasi

Evaluasi dilakukan dengan uji coba fungsional terhadap layanan yang disediakan oleh sistem satu per satu dan uji coba terhadap tingkat akurasi klasifikasi yang dihasilkan oleh sistem. Uji coba fungsional terhadap layanan yang disediakan oleh sistem memiliki tujuan untuk mengetahui hasil dari setiap proses dalam klasifikasi telah sesuai dengan analisis dan perancangan yang dilakukan. Uji coba terhadap tingkat akurasi klasifikasi yang dihasilkan bertujuan untuk mengetahui nilai *accuracy* dan *precision* klasifikasi menggunakan algoritma SVM berdasarkan kesesuaian hasil klasifikasi dan label.

- a. Pengujian akurasi sistem Pengujian akurasi sistem dilakukan dengan cara menghitung nilai dari *accuracy* dan *precision*. Rumus untuk menghitung nilai *accuracy* sebagai berikut.

$$\text{Acc} = \frac{\sum p}{n} \cdot 100$$

Keterangan:

v : Jumlah data benar

n : Jumlah dokumen

Sedangkan untuk menghitung *precision* menggunakan rumus sebagai berikut:

$$\text{Prec} = \frac{\sum vP}{nP} * 100$$

Keterangan:

vP : Jumlah data positif, negatif atau netral benar

nP : Jumlah dokumen positif, negatif atau netral

Pengujian dilakukan dengan jumlah data training yang berbeda. Pada setiap data training akan dilakukan 5 kali pengujian, sehingga tingkat akurasi sistem akan dihitung berdasarkan rata-rata dari setiap jumlah data training.

Data latih atau data training digunakan algoritma Support Vector Machine untuk mempelajari pola data berdasarkan ciri-ciri pada masing-masing kelas. Hasil pembelajaran tersebut kemudian diuji menggunakan data uji untuk mengetahui tingkat akurasi dalam memprediksi kelas pada data baru, proses ini disebut sebagai machine learning. Disini penulis menggunakan 1027 dataset komentar positif, negatif dan netral facebook dan 1000 dataset komentar positif, negatif dan netral pada instagram yang dibagi menjadi 5 (lima) kali pengujian dengan hasil akurasi klasifikasi algoritma Support Vector Machine menggunakan 60%, 70%, 80% dan 90% data training. Berikut di bawah ini skenario uji coba sistem untuk menentukan akurasi berdasarkan data latih dan data uji.

a. Skenario 1 dengan jumlah data latih 60%

Tabel 4.6 Skenario 1 dengan Data Latih 60%

Facebook dengan 1027 komentar					Instagram dengan 1000 komentar				
P	Acc	Prec 1	Prec -1	Prec 0	P	Acc	Prec 1	Prec -1	Prec 0
1	78.35	80.77	88.89	69.57	1	80.46	82.95	91.29	71.44
2	76.89	80.32	93.10	66.92	2	78.96	82.49	95.62	68.72
3	74.94	76.87	80.00	69.03	3	76.96	78.94	82.16	70.89
4	76.16	82.90	66.67	68.06	4	78.21	85.13	68.47	69.89
5	73.97	79.01	80.00	62.10	5	75.96	81.14	82.16	63.77
Rata-rata	76.06	79.97	81.73	67.13	Rata-rata	78.11	82.13	83.94	68.94

Keterangan:

- P : Pengujian ke-
 Acc : Accuracy
 Prec1 : Presisi Positif
 Prec-1 : Presisi Negatif
 Prec0 : Presisi Netral

Pada tabel 4.5 dilakukan 5 kali pengujian skenario 1 dengan jumlah data latih 60% mengacu pada pembagian data latih dan data uji pada tabel 4,3 diketahui bahwa rata-rata nilai akurasi facebook yaitu 76,06 bernilai lebih rendah daripada instagram 78,51. Demikian pula dengan nilai rata-rata presisi positif, negatif dan netral, instagram memiliki nilai lebih tinggi. Sedangkan rata-rata nilai presisi terbesar terdapat pada presisi komentar negatif yaitu 81,73 untuk facebook dan 83,94 untuk instagram.

b. Skenario 2 dengan jumlah data latih 70%

Tabel 4.7 Skenario 2 dengan Data Latih 70%

Facebook dengan 1027 komentar					Instagram dengan 1000 komentar				
P	Acc	Prec 1	Prec -1	Prec 0	P	Acc	Prec 1	Prec -1	Prec 0
1	76.70	78.68	88.24	66.67	1	78.77	80.80	90.62	68.47
2	75.08	77.01	86.21	67.74	2	77.11	79.08	88.53	69.57
3	77.35	79.55	96.67	67.96	3	79.43	81.69	99.28	69.80

Tabel 4.7 Skenario 2 dengan Data Latih 70% (Lanjutan)

4	76.70	80.98	79.11	68.32	4	78.77	83.16	81.24	70.16
5	78.32	77.35	86.84	76.67	5	80.43	79.44	89.19	78.74
Rata-rata	76.83	78.71	87.41	69.47	Rata-rata	78.90	80.84	89.77	71.35

Keterangan:

- P : Pengujian ke-
 Acc : Accuracy
 Prec I : Presisi Positif
 Prec -1 : Presisi Negatif
 Prec 0 : Presisi Netral

Pada skenario 2 percobaan dilakukan 5 kali dengan jumlah data latih 70% dari jumlah 1027 data set facebook dan 1000 dataset instagram diketahui nilai rata-rata akurasi dan presisi meningkat dari skenario 1. Pada akurasi terdapat kenaikan baik pada facebook maupun instagram yaitu sebesar 76,83 untuk facebook dan 78,90 untuk instagram. Nilai rata-rata presisi juga masih pada presisi negatif untuk nilai tertinggi yaitu 87,41 untuk facebook dan 89,77 untuk instagram.

c. Skenario 3 dengan jumlah data latih 80%

Tabel 4.8 skenario 3 dengan data latih 80%

Facebook dengan 1027 komentar

Instagram dengan 1000 komentar

P	Acc	Prec 1	Prec -1	Prec 0
1	82.52	89.74	83.33	70.42
2	83.01	88.10	85.71	71.19
3	81.07	83.59	78.95	76.27
4	82.04	80.00	100.00	82.81
5	76.70	78.69	88.89	69.70
Rata-rata	81.07	84.02	87.38	74.08

P	Acc	Prec 1	Prec -1	Prec 0
1	84.75	92.17	85.58	72.32
2	85.25	90.47	88.03	73.11
3	83.26	85.85	81.08	78.33
4	84.25	82.16	102.70	85.05
5	78.77	80.81	91.29	71.58
Rata-rata	83.26	86.29	89.74	76.08

Keterangan:

- P : Pengujian ke-
 Acc : Accuracy
 Prec1 : Presisi Positif
 Prec-1 : Presisi Negatif
 Prec0 : Presisi Netral

Pada skenario 3 percobaan dilakukan 5 kali dengan jumlah data latih 80% dari 1027 dataset komentar facebook dan 1000 dataset komentar instagram terdapat kenaikan baik akurasi maupun presisi dari skenario sebelumnya. Pada skenario ini, rata-rata nilai akurasi tertinggi masih terdapat pada instagram dengan nilai 83,26 daripada facebook yang hanya 81,07. Untuk nilai presisi baik facebook maupun instagram masih didominasi nilai rata-rata tertinggi presisi komentar negatif oleh instagram dengan nilai 89,74 daripada facebook dengan nilai 87,38

d. Skenario 4 dengan jumlah data latih 90%

Tabel 4.9 skenario 4 dengan data latih 90%

Facebook dengan 1027 komentar

Instagram dengan 1000 komentar

P	Acc	Prec 1	Prec -1	Prec 0
1	81.55	82.81	100.00	72.41
2	86.41	86.89	91.67	83.33
3	78.64	81.43	88.89	66.67
4	89.32	93.75	87.50	78.26
5	80.58	82.81	100.00	70.00
Rata-rata	83.30	85.54	93.61	74.14

P	Acc	Prec 1	Prec -1	Prec 0
1	83.75	85.05	102.70	74.37
2	88.74	89.23	94.14	85.58
3	80.76	83.63	91.29	68.47
4	91.73	96.28	89.86	80.37
5	82.76	85.05	102.70	71.89
Rata-rata	85.55	87.85	96.14	76.14

Keterangan:

- P : Pengujian ke-
 Acc : Accuracy
 Prec1 : Presisi Positif
 Prec-1 : Presisi Negatif
 Prec0 : Presisi Netral

Pada skenario 4 dengan data latih 90%, dilakukan 5 kali pengujian terdapat peningkatan rata-rata akurasi baik facebook maupun instagram yaitu 83,30

untuk facebook dan 85,55. Pada skenario ini memiliki nilai akurasi rata-rata tertinggi daripada skenario-skenario sebelumnya baik pada facebook maupun instagram. Untuk presisi masih didominasi rata-rata presisi negative dengan nilai 93,61 untuk facebook dan 96,14 untuk instagram.

3. Evaluasi Skenario pada Facebook dan Instagram

Diagram 4.1 Evaluasi Facebook



Pada diagram 4.1 untuk evaluasi facebook diketahui bahwa semakin banyak data latih yang digunakan, nilai rata-rata akurasi semakin tinggi yaitu dari 60%, 70%, 80% dan 90% memiliki nilai rata-rata akurasi 76,06; 76,83; 81,07 dan 83,30. Sedangkan nilai rata-rata presisi positif naik turun yaitu pada data latih 60%, 70%, 80% dan 90% memiliki nilai rata-rata presisi positif 79,97; 78,71; 84,02 dan 85,54. Dan nilai rata-rata presisi negatif juga naik turun yaitu pada data latih 60%, 70%, 80% dan 90% memiliki nilai rata-rata presisi negatif 81,73; 87,41; 87,38 dan 93,61. Berbeda dengan presisi positif dan negatif, presisi netral pada hasil rata-rata presisi di facebook memiliki kenaikan di setiap skenario data latih yang digunakan yaitu

pada data latih 60%, 70%, 80% dan 90% memiliki nilai rata-rata presisi netral 67,13; 69,47; 74,08 dan 74,14.

Diagram 4.2 Evaluasi Instagram



Pada diagram 4.2 evaluasi instagram dengan jumlah data latih 60%, 70%, 80% dan 90% juga memiliki kenaikan rata-rata akurasi dalam setiap penambahan data latih yang digunakan yaitu 78,11; 78,90; 83,26 dan 87,85. Sedangkan untuk presisi positif naik turun yaitu pada data latih 60%, 70%, 80% dan 90% memiliki nilai rata-rata presisi positif 82,13; 80,84; 86,29 dan 87,85. Dan untuk presisi negatif juga mengalami naik turun yaitu pada data latih 60%, 70%, 80% dan 90% memiliki nilai presisi negatif sebesar 83,94; 89,77; 89,74 dan 96,14. Berbeda dengan presisi positif dan negative, sama halnya pada facebook presisi netral di instagram juga mengalami kenaikan yaitu pada data latih 60%, 70%, 80% dan 90% dengan nilai presisi netral 68,94; 71,35, 76,08 dan 76,14.

4. Evaluasi skenario dengan Count Vector dan TF IDF

Tabel 4.10

Pengujian Count Vector dan TF IDF pada Facebook

Nama	Komentar	Akurasi	precision	Recall	F1 score
Count vector	Positif		59%	79%	68%
	Negatif	57%	54%	48%	47%
	Netral		67%	67%	67%
TF-IDF Word Level	Positif		55%	91%	68%
	Negatif	56%	82%	25%	37%
	Netral		100%	17%	29%
TF-IDF Ngram level	Positif		53%	100%	69%
	Negatif	57%	100%	33%	42%
	Netral		100%	50%	67%
TF-IDF Char level	Positif		55%	93%	69%
	Negatif	56%	88%	20%	28%
	Netral		100%	17%	29%

Pada tabel 4.10 dapat diketahui bahwa penghitungan pada media sosial facebook dengan ekstraksi Count Vector terdapat nilai akurasi 57% dengan tingkat presisi komentar positif, negative dan netral adalah 59%, 54% dan 67%. Nilai recall 79%, 48% dan 67% serta F1 score 68%, 47% dan 67%. Sedangkan dengan TF IDF Word Level nilai akurasi 56% dengan tingkat presisi komentar positif, negative dan netral adalah 55%, 82% dan 100%. Nilai recall 91%, 25% dan 17% serta F1 score 68%,

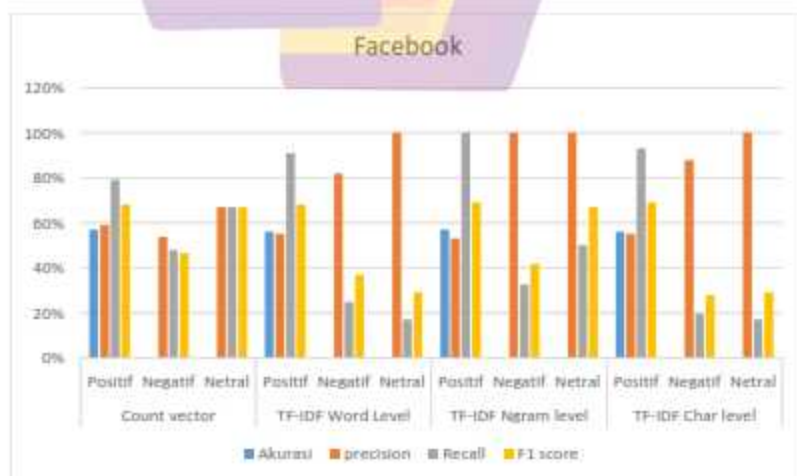
37% dan 29%. Dengan TF-IDF N-Gram Level memiliki nilai akurasi 57% dengan tingkat presisi komentar positif, negative dan netral adalah 33%, 100% dan 100%. Nilai recall 100%, 33% dan 50% serta F1 score 69%, 42% dan 67%. Serta nilai akurasi pada TF-IDF Char Level 56% dengan tingkat presisi komentar positif, negative dan netral adalah 55%, 88% dan 100%. Nilai recall 93%, 20% dan 17% serta F1 score 69%, 28% dan 29%.

Tabel 4.11
Pengujian Count Vector dan TF IDF pada Instagram

Nama	Komentar	Akurasi	precision	Recall	F1 score
Count vector	Positif		64%	71%	68%
	Negatif	59%	54%	46%	47%
	Netral		67%	67%	67%
TF-IDF Word Level	Positif		54%	90%	54%
	Negatif	57%	82%	25%	37%
	Netral		97%	17%	30%
TF-IDF Ngram level	Positif		53%	87%	69%
	Negatif	59%	89%	31%	42%
	Netral		90%	56%	67%
TF-IDF Char level	Positif		47%	93%	69%
	Negatif	57%	76%	23%	32%
	Netral		98%	17%	21%

Pada tabel 4.11 dapat diketahui bahwa penghitungan pada media sosial instagram dengan ekstraksi Count Vector terdapat nilai akurasi 59% dengan tingkat presisi komentar positif, negative dan netral adalah 64%, 54% dan 67%. Nilai recall 71%, 46% dan 67% serta F1 score 68%, 47% dan 67%. Sedangkan dengan TF IDF Word Level nilai akurasi 57% dengan tingkat presisi komentar positif, negative dan netral adalah 54%, 82% dan 97%. Nilai recall 90%, 25% dan 17% serta F1 score 54%, 37% dan 30%. Dengan TF-IDF N-Gram Level memiliki nilai akurasi 59% dengan tingkat presisi komentar positif, negative dan netral adalah 53%, 89% dan 90%. Nilai recall 87%, 31% dan 56% serta F1 score 69%, 42% dan 67%. Serta nilai akurasi pada TF-IDF Char Level 57% dengan tingkat presisi komentar positif, negative dan netral adalah 47%, 76% dan 98%. Nilai recall 93%, 23% dan 17% serta F1 score 69%, 32% dan 21%.

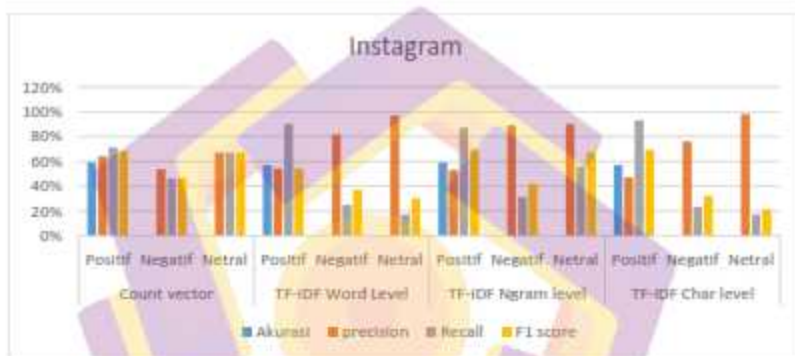
Diagram 4.3
Perbandingan Nilai Akurasi dengan Fitur Count Vector dan TF-IDF
pada Facebook



Pada diagram 4.3 pada media facebook dapat dilihat bahwa nilai akurasi tertinggi ada pada Count Vector sedangkan nilai presisi tertinggi ada pada TF-IDF baik word level, count level dan char level.

Diagram 4.4

Perbandingan Nilai Akurasi dengan Fitur Count Vector dan TF-IDF
pada Instagram



Pada diagram 4.4 pada media instagram sebagaimana pada facebook dapat dilihat bahwa nilai akurasi tertinggi ada pada Count Vector sedangkan nilai presisi tertinggi ada pada TF-IDF baik word level, count level dan char level.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan yang dihasilkan pada uji coba dan evaluasi penelitian ini diantaranya adalah:

- a. Data mentah hasil scraping data pada facebook dan instagram AG243 dengan kata kunci covid-19, covid19, covid, corona, dan pandemi dapat diketahui bahwa jumlah komentar pada facebook lebih banyak sejumlah 1959 daripada instagram sejumlah 1575 komentar. Sedangkan kata kunci yang dominan pada facebook menggunakan "covid" sedangkan instagram dengan kata kunci "pandemi".
- b. Opini masyarakat Kediri di media sosial terhadap penyebaran informasi Covid-19 pada akun media sosial AG243 (radio andika) dapat diketahui berdasarkan tingkat akurasi baik pada facebook maupun instagram, masyarakat Kediri banyak berkomentar negatif yaitu dengan akurasi tertiggi 83,30 dengan jumlah data latih 924 dan data testing 103 untuk facebook dan 85,55 dengan jumlah data latih 900 dan data testing 100 untuk instagram.
- c. Jika berdasarkan pendekatan ekstraksi fitur SVM dengan count vector, tf-idf word level, tf-idf ngram level dan tf-idf char level. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur count vector dan tf-idf ngram level dengan nilai 57% pada facebook dan 59% pada instagram.

- d. Support Vector Machine belum bisa digunakan sebagai algoritma yang menangkap urutan karena kata yang dibolak balik meskipun maknanya berbeda tetap diakui sama oleh mesin dengan klasifikasi SVM, hal ini juga terbukti dengan jumlah akurasi yang sangat kecil.

5.2. Saran

Terdapat beberapa saran untuk penelitian selanjutnya, diantaranya :

- a. Karena SVM (Support Vector Machine) belum bisa menangkap urutan kata dalam kalimat maka diperlukan algoritma klasifikasi yang bisa menangkap urutan kata dengan makna yang sesuai misalnya dengan algoritma HRRN (Highest Response Ratio Next) yang memperhatikan rasio dengan rasio respon paling tinggi atau LSTM (Long Short-Term Memory).
- b. Dapat dikembangkan peluang penelitian pada akun institusi lain dan pada media sosial yang berbeda.
- c. Dapat digunakan pendekatan lain misalkan dengan Lesicon Based, word2vec atau yang lainnya

DAFTAR PUSTAKA

PUSTAKA BUKU

- Abram, C. 2016. "Facebook For Dummies": Edition 6(Vol. 6). John Wiley & Sons.
- Faisal, M. R. 2017. "Seri Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R. Indonesia". Net Developer Community.
- Feldman, R., & Sanger, J. 2007. "The Text Mining Handbook Advanced Approaches In Analyzing Unstructured Data." New York : Cambridge University Press. (Jamil, 2017).
- Kusrini. 2006. "Sistem Pakar Teori dan Aplikasi". Andi Offset
- 2008. "Aplikasi sistem Pakar Lanjut". Andi Offset
- Laksana, E. A., & Sulianta, F. 2017. "Analisis Dan Studi Komparatif Algoritma Klasifikasi Genre Musik". 67-72
- Lee, L., & Pang, B. 2008. "Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval", 2(1-2): 1-135.
- Liu, B. 2012. "Sentiment Analysis and Subjectivity. Synthesis: Lectures on Human Language Technologies." USA: Morgan & Claypool Publishers.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. 2012. "Foundations of Machine Learning". MIT Press
- Prasetyo, Eko. 2014. "Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab". Yogyakarta: ANDI
- Sugiyono. 2015. Metode Penelitian Kuantitatif, Kualitatif, dan R&D. 22 ed. Bandung: Alfabeta.
- Suyanto. 2017. "Data Mining Untuk Klasifikasi dan Klusterisasi Data". Bandung: Informatika Bandung.
- Tan, P., Steinbach, M., & Karpapne, A. 2006. "Introduction To Data Mining". USA: Addison-Wesley.
- Turban, E., Aronson, J., & Liang, T. 2005. "Decision Support System And Intelligent System". Upper Saddle River, New Jersey USA: Prentice Hall.
- 2005. "Service Quality Satisfaction". Yogyakarta: Andi.

PUSTAKA MAJALAH, JURNAL ATAU PROSIDING

- Alhaji, Moehammad. 2020. "Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19". preprints
- Athira, Wanda. 2018. "Analisis Sentimen Cyberbullying Pada Komentar Instagram Dengan Metode Klasifikasi Support Vector Machine". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer
- Budi, Eko. 2019. "Analisis Sentimen Calon Presiden Indonesia 2019. Berdasarkan Komentar Publik Di Facebook". Jurnal Eksplora Informatika
- Burhanudin, Ikhsan. 2020. "Ancaman Krisis Ekonomi Global Dari Dampak Penyebaran Virus Corona (Covid 19)". Akmen
- Fatmawati. 2017. "Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus: Akun Fb Group Iraise Helpdesk)". Jurnal CoreIT
- Fauzi, Akhmad. 2019. "Sentimen Analisis Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Bayes". Jurnal Informatika
- Janoušek, V., Moyen, J-F., Martin, H., Erban, V., & Farrow, C. 2015. "Geochemical Modelling of Igneous Processes—Principles And Recipes in R Language: Bringing the Power of R to a Geochemical Community". Springer Geochemistry
- Lappeman, James. 2020. "Studying Social Media Sentiment Using Human Validated". Elsevier
- Maulana, Akbar. 2018. "Analisis Sentimen Media Sosial Univ.Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM ". Seminar Nasional Teknologi Informasi dan Multimedia
- Murty, M., & Raghava, R. 2016. "Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks.Springer".
- Puspita. 2014. "Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi SD Di Kab. Magelang". Jurnal Gaussian
- Putri, D.U.K. 2016. "Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten". Program Studi Ilmu Komputer FMIPA UGM Yogyakarta.

Salam, Abu. 2018. "Analisis Sentimen Data Komentar Sosial Media Fb Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)". Prosiding Sintak

Saraswati, N.S. 2011. "Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis." Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.

PUSTAKA ELEKTRONIK

<https://corona.kedirikota.go.id/> diakses 12 April 2020

<http://covid19.kedirikab.go.id/> diakses 12 April 2020

<http://infocovid19.jatimprov.go.id/> diakses 12 April 2020

<https://inacovid19.maps.arcgis.com/apps/opsdashboard/index.html#/4411f5e9c69d4ca4be31ac805a0267be> diakses 12 April 2020

